

УДК 612.78

П'єх О. – ст. гр. ПМЗпм-71

*Тернопільський державний технічний університет імені Івана Пулюя***ПІДВИЩЕННЯ ЯКОСТІ РОЗПІЗНАВАННЯ МОВНОГО СИГНАЛУ
ЗА ДОПОМОГОЮ ЛІНІЙНОЇ СМУГОВОЇ ФІЛЬТРАЦІЇ ЙОГО
ЛОГАРИФМІЧНОГО АМПЛІТУДНОГО СПЕКТРУ**

Науковий керівник: к.м.н., доцент Гевко О.В.

В основі сучасних систем розпізнавання мови використовується процедура порівняння поточного спектру мовного сигналу із спектральними еталонами, отриманими в процесі навчання системи (методи прихованих Марківських моделей, динамічного програмування, нейромережевий метод). У таких випадках отримуємо порівняно чітке розпізнавання для фіксованого диктора із стабільним вимовою і незмінних характеристик навколишнього середовища. При порушенні цих умов якості розпізнавання різко знижується. Причиною цього є варіабельність вимови диктора, вплив шумів і частотних спотворень, зумовлених акустикою приміщення, характеристиками мікрофона і каналу зв'язку. Перераховані чинники призводять до значних варіацій спектра мовного сигналу, що ускладнює процес порівняння з наявними спектральними еталонами, в результаті чого значно погіршується якість розпізнавання. Тому, розроблення методів покращення розпізнавання мови за наявності частотних спотворень і адитивних завад є актуальним.

Пропонується використати лінійну смугову фільтрацію логарифмічного амплітудного спектру мовного сигналу з подальшим нелінійним перетворенням, що моделює ефект латерального гальмування в слуховому аналізаторі.

Подано лінійну модель мовного сигналу в частотній області у вигляді[1]:

$$S(z) = H(z)E(z)W(z) \quad (1)$$

де $H(z)$ - частотна характеристика мовного тракту, $E(z)$ - спектр шумового або імпульсного сигналу, $W(z)$ - характеристика фільтру, що описує частотні спотворення мовного сигналу.

Після логарифмування рівняння (1) набуде вигляду:

$$F(z) = \lg S(z) = \lg H(z) + \lg E(z) + \lg W(z) \quad (2)$$

Розкладемо спектр $F(z)$ на складові за допомогою лінійної фільтрації. Потім, шляхом смугової фільтрації логарифмічного спектру $S(z)$ зменшимо вплив складових, які обумовлюють варіації спектру мовного сигналу. Для стабілізації $F(z)$ здійснюємо нелінійне перетворення, за допомогою якого в спектрі зберігаються фрагменти, пов'язані з максимумами (формантами) $H(z)$, за наявності фонових широкосмугових шумів.

Таким чином, полегшується процес порівняння досліджуваного сигналу з еталонами і підвищується якість розпізнавання мовного сигналу.

Література:

1. Венцов А.В. Касевич В.Б. Современные модели восприятия речи: критический обзор.— Издательство Санкт-Петербургского университета. 1994.